P2P 重叠网的立体化研究与设计①

徐海斌，张云华，张烨飞，朱 光

(浙江理工大学 信电学院，杭州 310018)

摘 要：在分析研究 DHT 算法与 B+树模型的基础上提出一种新的网络模型——圆台网络，该圆台模型将 DHT 网络由二维带到三维，由平面转成了立体，通过把节点的路由表设成可动态增长的二维表，现实了节点的立体查寻，提高查询效率。有限长度的后继列表的更新及有限的传递次数，确保了整个网络的稳定畅通，在确保网路畅通的同时也使得系统在维护开销上比 chord 算法大为降低。

关键字：P2P 网络；圆台网络；路由；父节点；查寻；维护开销

Building Three-Dimensional Network for P2P Overlay Network

XU Hai-Bin, ZHANG Yun-Hua, ZHANG Yie-Fei, ZHU Guang

(Information and Electronic College, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: Analyzing the algorithm of DHT and studying the thinking of the B + tree, we got a new network models -Frustum of a cone network. This network models will bring us to three-dimensional network from two-dimensional network, from plane to solid. Setting the end points, the routing table to be dynamic growth two-dimensional tables, we realized three-dimensional end point search and improved query efficiency. The updates of Limited length of follow-up list and limited delivery times, ensure that the entire network will be stable, also ensure that the system be much less than the chord algorithm in costs of the maintenance.

Keywords: P2P network; frustum of a cone network; routing; parent node; search; costs of maintenance

1 概述

通过某种方式的查询实现资源的定位是 P2P 网络的一个核心问题。最初的 P2P 系统，例如，Napster 利用中心索引服务器存储数据的元数据信息，中心索引服务器很容易成为系统性能的瓶颈和不稳定的根源。其后的一些系统，如 Gnutella 和 Freenet，由于没有中央服务器，在搜寻数据时是以 flooding 的方式将消息散布在网络上，因此会造成消息泛滥的问题。

目前，基于分布式哈希表(Distributed Hash Table,DHT)的模型很好地解决了前期 P2P 存在的单点崩溃、泛洪消息泛滥、可扩展性差等问题，如 Chord[1],CAN,Pastry[2]和 Tapestry。其中，Chord 具有简单、可靠、查找高效等其他 DHT 模型所不具备的特点。然而 Chord 仍存在不足，可进一步改进优化其性能。研究表明，Chord 算法的性能在很大程度上取决于用于路由的 Fingertable，而且 chord 算法为维护网络稳定性所作出的维护开销仍然很大，而且没考虑节点的异构性，维护机制复杂、开销大，Chord 路由表中存在严重信息冗余。研究人员对 Chord 进行很多研究，并试图对其改进[3]。

本文对 DHT 算法进行深入分析并参考了硬盘数据组织 B+树模型，通过改变网络平面结构设计了一个立体网络模型-圆台网络模型。该模型变平面网络为立体网络，让查寻可横向、纵向查寻，减少了查寻转发次数，缩短搜索路径的长度，缩小搜索延迟，提高搜索效率，同时该算也大大降低网络维护开销。

2 圆台网络模型

节点之间所组成的网络模型如图 1。

若网络共拥有：n^h 个节点，每个节点都拥有且维护 h*n 个项的后继路由表，该表分 h 行，每行 n 个节点信息，第一行是顶层所有 n 个元素路由信息，第二，

① 收稿时间:2010-08-02;收到修改稿时间:2010-09-14

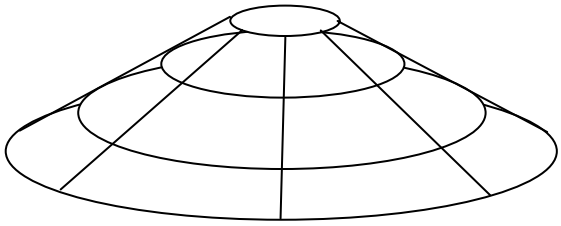

图 1　圆台网络模型图(Frustum of a cone network）

第三到第 h 行分别是该节点在所在环中 n 个连续后继节点的路由信息。为方便理解将 n 取 10，h 取 4，图 1 就是一个拥有 10000 个节点的网络。最底层是所有节点巩筑的圆环，有 10000 个节点的网络，该网络共分四层，从上往下，第一层节点的编号是：0.1，0.2，……，0.9，1 共 10 个，第二层的节点编号是：0.01，0.02，……，0.99，1 共 100 个，第三层的节点编号是：0.001，0.002，……，0.999，1 共 1000 个，第四层节点的编号是：0.0001，0.0002，……，0.9999，1 共 10000 个。该网络，从最顶层开始，每个节点可最多自分裂成 10 个子节点，到第二层网络便容纳下 100 个节点，逐级向下扩展，当达到第 h 层时，网络就可容纳下个节点。

以下表 1，表 2 是两个 h 为 4 的分别编号为 0 和编号为 0.6625 节点的圆台网络的路由表(以下图表为简化表达，只列出了相应的节点号，省去节点对应网络路由信息项)。

表 1　号节点路由表（0 node routing table)

后继队首	2 号	3 号	4 号	5 号	6 号	7 号	8 号	9 号	尾
0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009	0.01
0.0001	0.0002	0.0003	0.0004	0.0005	0.0006	0.0007	0.0008	0.0009	0.001

表 2　0.6625 号节点路由表(0.6625 node routing table)

后继队首	2 号	3 号	4 号	5 号	6 号	7 号	8 号	9 号	尾
0.7	0.8	0.9	1	0.1	0.2	0.3	0.4	0.5	0.6
0.67	0.68	0.69	0.7	0.71	0.72	0.73	0.74	0.75	0.76
0.663	0.664	0.665	0.666	0.667	0.668	0.669	0.67	0.671	0.672
0.6626	0.6627	0.6628	0.6629	0.663	0.6631	0.6632	0.6633	0.6634	0.6635

该路由表最底行保存其后继的 10 个节点的路由信息，上一行是其父节点及其父节点的 9 个后继，第二行是第三行的父节点及 9 个后继节点，以此类向上推，当然最顶行记录所有 10 个顶层节点的路由信息。

在圆台网络每个节点都拥有这样一份路由表，通过该路由表可以很快找到所要查寻的节点信息。比如：0.6625 号节点要查寻 0.3548 号节点的信息，首先，在自身路由表的第一行首先找到小于 0.3548 且最接近的数值即 0.3 号节点，接着转到 0.3 号节点的路由表在其第二行找到 0.35 号节点，然后转到 0.35 号节点在路由表的第三行找到 0.354 号节点，最后在 0.354 号节点路由表的第 4 行找到 0.3548 号节点信息。

具有 1000 个节点的网络，n 取 10，请求转发将在 3 步完成，同样 1000000 个节点的网络，n 取 10，请求转发 6 步就可完成。对于一个具有 N 个节点的圆台网络，对其节点查询只须进行 $O(\log_n N)$个路由跳, 与 chord 算法的平均查找需要$(\log_2 N)/2$ 相比，路由所产生的跳数更少，查寻的速度更快。

3　网络稳定化

为确保网络的稳定，一个具有 N 个节点的网络，每个节点都保留了$(\log_n N)$*n 个后继节点的路由信息，其中最底层的 n 个后继节点路由信息，利用其状态变化的更新，确保了整个网络的连通与一致性。每个节点都会定期向其后继节点发送信息，检查后继节点路由信息状态，首先检查后继节点是否仍然存在，若超时无响应，则将其从队首删除，若存在，则查看其路由状态是否处在已更新状态，若状态已更新，则需要将后继后点及其 n-1 后继信息复制到自身对应的后继队列中。

节点状态的设置，由节点检查其自身路由信息有无变化及后继的 n-1 个后继的路由信息有无变化，若其中有一个节点已变化，则设置为更新状态。注意若是后继队列的第 n 个节点信息有变化，状态不用设为更新状态。因为该节点自身和后继的 n-1 节点组成其前趋的后继队列，若只是自身后继队尾有变化，则不影响前趋的后继队列表。所以一个节点发生变化，将导致其 n 个前趋的后继路由表需要更新。圆台网络一共有 $\log_n N$ 层，故一个节点的变化总共引发$(\log_n N)$*n 个节点的路由信息需要更新。

圆台网络路由维护的开销为：O(($\log_n N$)*n)，而 chord 网络所产生平均路由维护开销是：O(($\log_2 N$)*($\log_2 N$)),当网络节点数达到 10000 时，圆台网络 n 取 10，一个节点的变化总共引 $\log_{10}10000$*10=40 个节点路由信息更新，而 chord 算法需要($\log_2 10000$)*($\log_2 10000$)≈177 个节点的路由信息更新开销，随着节点的增多圆台网络的效率将更加明显。由此可见，圆台网络与 chord 相比，它可大幅降低路由维护开销。

4 节点到达

每个节点拥有一个 n 位置后继列表，同时也可由自分裂成 n 个孩子节点的位置，当有新节点申请加入网络时，被申请的节点首先检查自身节点路由表的最底层的后继队列是否有空位置，若有则直接插入到后继列入中，若不是队列的最后一个节点，把路由表状态设置成已更新状态，以引起其前趋节点的路由表的更新。若后继队列已满，则把申请加入的信息向后传递，直到有空位为止。若恰好，所有后继的都已满，此时网上的节点数恰好是 n^h 个，(若 n 取 10，则节点数恰好是 10，或 100，1000，……）,这时候，申请加入的信息通过网络被传递回来，此时圆台网络需要向下增加一层，即节点从自身分裂出 n 个孩子节点，路由表的行数增加一行。被申请加入节点成为新加入节点的直接后继，并更新路由表的状态，让被申请加入节点的前趋成为新加入节点的前趋。

随着网络规模的增大，若是始终由节点传递来寻找空位置，则很可能要让节点等待很长时间。为了不让申请节点等待过长时间，可以采用两种方案。

第一种是父节点状态标志法，每个父节点持有一个孩子节点是否已满的状态标志。从最顶层开始这样便可以在父节点或其祖先节点中先找到孩子节点未满的节点，再申请加入。

第二种是直接生成孩子节点法，只要自身后继队列没有空位置，就扩展路由表，向下生成 n 个节点的空位，新节点便可直接加入其中。

以后两种方法各有利弊，第一种方法它保持圆台模型的完美性，确保每层节点满个节点后再向扩展，但需要花一定的时间去查找空位。第二种方法，省去查找空位的时间，但会使网络形状看起有些参差不齐，但网络节点的申请分布总体均匀的话，它几乎不影响网络性能。

5 节点故障

当节点定期向后续节点申请路由状态时，后继节点长时间无响应，此时应考虑为节点故障，应将其从队列中删除，并更新路由表的状态。为不使数据失效，节点在活动状态时就把相关数据传给后继节点进行备份。

6 结束语

本文在分析研究 DHT 算法与 B+树模型的基础上提出一种新的网络模型-圆台网络，该圆台模型将 DHT 网络由二维带到三维，由平面转成了立体，通过把节点的路由表设成可动态增长的二维表，现实了节点的立体查找，提高查询效率。有限长度的后继列表的更新及有限的传递次数，确保了整个网络的稳定畅通，在确保网路畅通的同时也确保系统在维护开销上比 chord 算法大为降低。初步实验表明该网络模型较好改善了 P2P 网络的性能。

参考文献

1 王新生,梁平,张云超,王伟杰,丁学永.结构化 P2P 路由协议的改进.计算机工程,36,(10):106-107.

2 Stoica I, Morris R, Karger D, et al. Chord: a scalable peer-to-peer lookup service for internet applications. Proc. of ACM SIGCOMM'01. San Diego, California, 2001.

3 Row S. Pastry: Scalable, Distributed Object Location and Routing for Large-scale Peer-to-Peer Systems. [2008-06-21]. http://www.research.microsoft.com/~antr/PAST/pastry.ps

4 Steinmetz R, Wehrle K. 王玲芳,陈焱译.Peer-to-Peer Systems and Applications.北京:机械工业出版社,2008.